

Claims

What is claimed is:

- 5 Sub
al
1. A method for classifying data, comprising the steps of:
classifying objects in a domain dataset using a data classification model,
said data classification model having a bias;
evaluating the performance of said classifying step; and
modifying said bias based on said performance evaluation.
 - 10 2. The method of claim 1, wherein said steps of classifying and evaluating
are performed for a plurality of said domain datasets and wherein said method further
comprising the steps of recording a performance value for each combination of said
domain datasets and said bias.
 - 15 3. The method of claim 2, further comprising the step of processing said
recorded performance values for each combination of said domain datasets and said bias
to generate one or more rules, each of said rules specifying one or more characteristics of
said domain datasets and a corresponding bias that should be utilized in one of said data
classification models.
 - 20 4. The method of claim 3, further comprising the step of selecting a data
classification model for classifying a domain dataset by comparing characteristics of said
domain dataset to said rules.
 - 25 5. The method of claim 1, wherein said domain dataset is represented using a
set of meta-features.

6. The method of claim 5, wherein said meta-features includes a concept variation meta-feature.

7. The method of claim 5, wherein said meta-features includes an average weighted distance meta-feature that measures the density of the distribution of said at least one domain dataset.

8. A method for classifying data, comprising the steps of:
classifying objects in a plurality of domain datasets using one of a number
of data classification models, each of said data classification models having a
corresponding bias;
evaluating the performance of each of said domain dataset classifications;
maintaining a performance value for each combination of said domain
datasets and said bias;
processing said performance values for each combination of said domain
datasets and said bias to generate one or more rules, each of said rules specifying one or
more characteristics of said domain datasets and a corresponding bias that should be
utilized in one of said data classification models; and
selecting a data classification model for classifying a domain dataset by
comparing characteristics of said domain dataset to said rules.

9. The method of claim 8, further comprising the step of modifying at least one of said biases based on said performance evaluation.

10. The method of claim 8, wherein said domain dataset is represented using a set of meta-features.

11. The method of claim 10, wherein said meta-features includes a concept variation meta-feature.

12. The method of claim 10, wherein said meta-features includes an average weighted distance meta-feature that measures the density of the distribution of said at least one domain dataset.

13. A method for classifying data in a domain dataset, comprising:
applying an adaptive learning algorithm to said domain dataset to select a data classification model, said data classification model having a bias;
classifying objects in said domain dataset using said selected data classification model;
evaluating the performance of said classifying step;
maintaining an indication of said performance of said model for said domain dataset;
repeating said applying, classifying and evaluating steps for a plurality of said domain datasets; and
processing said performance values for each combination of said domain datasets and said bias to adjust one or more rules for subsequent data classification, each of said rules specifying one or more characteristics of said domain datasets and a corresponding bias that should be utilized in one of said data classification models.

14. The method of claim 13, further comprising the step of selecting a data classification model for classifying a domain dataset by comparing characteristics of said domain dataset to said rules.

15. The method of claim 13, further comprising the step of modifying at least one of said biases based on said performance evaluation.

16. A system for classifying data, comprising:
a memory that stores computer-readable code; and
a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:
classify objects in a domain dataset using a data classification model, said data classification model having a bias;
evaluate the performance of said classifying step; and
modify said bias based on said performance evaluation.

17. The system of claim 16, wherein said processor is further configured to classify said objects and evaluate said performance for a plurality of said domain datasets and wherein said processor records a performance value for each combination of said domain datasets and said bias.

18. The system of claim 17, wherein said processor is further configured to process said recorded performance values for each combination of said domain datasets and said bias to generate one or more rules, each of said rules specifying one or more characteristics of said domain datasets and a corresponding bias that should be utilized in one of said data classification models.

19. The system of claim 18, wherein said processor is further configured to select a data classification model for classifying a domain dataset by comparing characteristics of said domain dataset to said rules.

20. The system of claim 16, wherein said domain dataset is represented using a set of meta-features.

21. A system for classifying data, comprising:

a memory that stores computer-readable code; and
a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

classify objects in a plurality of domain datasets using one of a number of data classification models, each of said data classification models having a corresponding bias;

evaluate the performance of each of said domain dataset classifications;
maintaining a performance value for each combination of said domain datasets and said bias;

process said performance values for each combination of said domain datasets and said bias to generate one or more rules, each of said rules specifying one or more characteristics of said domain datasets and a corresponding bias that should be utilized in one of said data classification models; and

select a data classification model for classifying a domain dataset by comparing characteristics of said domain dataset to said rules.

22. An article of manufacture for classifying data, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to classify objects in a domain dataset using a data classification model, said data classification model having a bias;

a step to evaluate the performance of said classifying step; and

a step to modify said bias based on said performance evaluation.

23. An article of manufacture for classifying data, comprising:
a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:
5 a step to classify objects in a plurality of domain datasets using one of a number of data classification models, each of said data classification models having a corresponding bias;
a step to evaluate the performance of each of said domain dataset classifications;
10 a step to maintaining a performance value for each combination of said domain datasets and said bias;
a step to process said performance values for each combination of said domain datasets and said bias to generate one or more rules, each of said rules specifying one or more characteristics of said domain datasets and a corresponding bias that should
15 be utilized in one of said data classification models; and
a step to select a data classification model for classifying a domain dataset by comparing characteristics of said domain dataset to said rules.